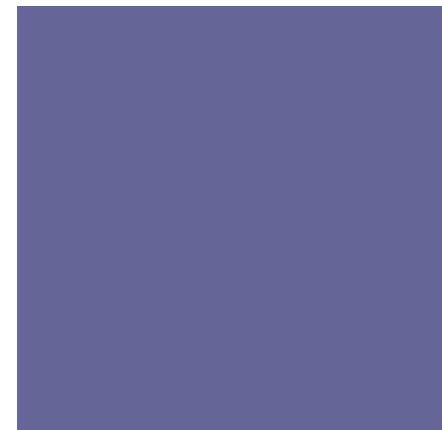




**Using Web archives for
researchers :
experience,
challenges,
issues**
Valérie Schafer



WWW

- 1- When
- 2- What
- 3- Why

Conference Saving the Web
Brussels, 18 October 2019





When did I come to Web Archives?



- I started by studying the Web without Web Archives
 - 2011 = discovery of Internet Archive thanks to Niels Brügger
 - The French WEB90 project (2014-2018)
 - ASAP Project = a turn to DSN like Twitter (2016)
 - *Negotiating the Web of the Past* (paper, 2016); *Do Web Archives have Politics ?* (presentation, 2017); *Qu'est-ce qu'une archive de Web?* (collective book, 2019)
- Through WARCnet (funded by the Danish Research Council) : comparing Web domains / analysing transnational events (Terrorist attacks, Refugee crisis, EU parliamentary elections, Olympic Games in London...)

Web



+ What do we do with Web Archives ?

- Study of Web archiving processes and heritagization of our digital past
- Web archives as illustrations
- Experience of the Web of the past / « re-enactment »
- Getting some missing information to rebuild the digital landscape
- Crossing Web archives with other sources (printed, digitized, born-digital...)



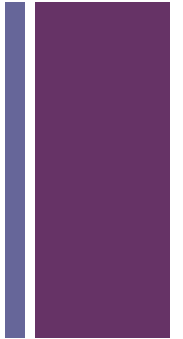
✚ Qualitative studies:

Focused studies on some webpages;

diachronic study of websites;

Mapping networks and

Web landscape on a precise topic ...



Les lieux de savoir

Dans les généalogies nourrissant l'histoire d'Internet, le rôle du monde scientifique depuis les années 1970 est crucial : il fournit les premiers développeurs et utilisateurs qui peuplent les listes et groupes de discussion, faisant d'Internet un grand forum savant. L'appropriation du Web par les professionnels du savoir relève alors d'une forme de migration innovante : les anciens outils aux interfaces rudimentaires ne sont pas délaissés, mais les nouveaux, au potentiel graphique et interactif riche, sont vite expérimentés par ces adoptants précoces. Les écoles, universités et bibliothèques se dotent rapidement de pages Web. De plus, le champ des utilisateurs savants s'élargit : depuis son centre académique, gagnant une population étudiante, il s'étend progressivement à sa périphérie, via les actions de médiation des associations et des professionnels. Le Web accompagne l'ouverture des lieux de savoir à la société.
Camille Paloque-Berges



Unité réseaux du CNRS

<http://www.urec.fr>

Le CNRS est l'une des premières institutions scientifiques à mettre à disposition de son public de chercheurs une boîte à outils pour la découverte et l'appropriation du Web. D'abord généraliste, l'UREC devient en 1999 un annuaire spécialisé des lieux de l'enseignement supérieur et de la recherche.

Archive du 16 mai 1997 à 02:58 » Voir d'autres dates



Biblio Online

<http://www.bol.ocd.fr>

Portail vers le monde des bibliothèques, Biblio Online est créé en 1997 par des développeurs de logiciels de gestion pour les bibliothécaires. À destination des professionnels et du grand public, il exemplifie les ponts entre traditions savantes et innovations technologiques.

Archive du 16 avril 1997 à 04:59 » Voir d'autres dates

• לא רואים עברית | הפוך לדף הבית | תמיכה ושירות | חלון החדשות של הארץ | **Haaretz English Edition** haaretz.com

HAARETZ Daily Newspaper Ltd. | **הוצאת עיתון הארץ בע"מ**

למנוי על העיתון המודפס

haaretz.co.il

הארץ Online

שען ישראל: 01.37 (GMT+3)

יום חמישי כ"א בסיון תשס"ז, 10 ביוני 2004

חדשות | כלכלה | גלריה | ספורט | העיתון המודפס | מדריך טלוויזיה | ארכיון הארץ | **קפטן אינטרנט**

הארץ ספרים

מיוחד לשבוע הספר

סיבה בלילה - אובל

לפגוע ודביות

סיבה סודית - סגב האסטרדה

בזכות הארץ: סגב החיים

עשרת השנים המפילות בארץ

חילת הארץ - סגב הנספר

מבזקים
אחרים

00:16 הארץ

ר"ם טוקיה ארדואן מתה בישראל על ישראל במיטות מדיות הג'8, ואם כי פשיטת ישראל בלבטון ומפנה מדיות על חוסר כנות בהצגת השום (בתק גיטמן)

המשטרה: עלייה של כ-15% באלימות של מתנחלים נגד פלשתינאים מתחילת השנה

על פי נתונים, הפלייה חלה בעיקר בשומרון ובתברון. ברוב המקרים מדובר בגניבת ציוד, השמדת רכוש ותקיפת עוברי אורח. 5 תושבי יצנה נעצרו בחשד שהתקפו רוצח פלשתינאי



The virtual exhibit by Anat Ben-David
20 yrs of Israeli Websites

<https://www.anatbendavid.info/single-post/2018/04/25/Virtual-Exhibition-20-years-of-Israeli-Websites>



■ Trying to get a whole picture of the Web of the Past



CURRENT PROJECTS

[Probing a Nation's Web Domain](#)

RESAW –
Transnational
Research
Infrastructure

Fundamental Tools
for Web Research

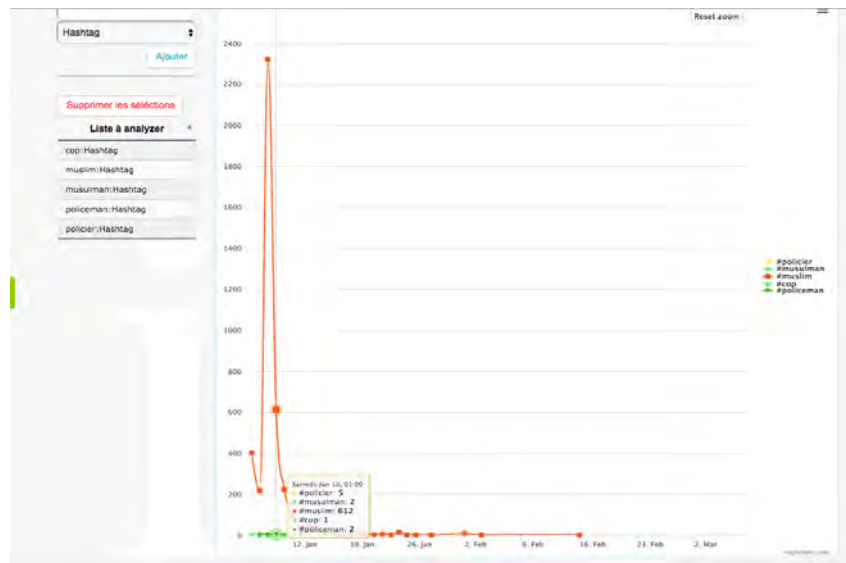
Probing a Nation's Web Domain – the Historical Development of the Danish Web

Purpose: *The project aims at a) mapping the development of the Danish web domain .dk, b) developing the research infrastructure needed to conduct the study, c) become a blueprint for a European comparative study.*

Expected outcomes: *The outcome of the project will be: a) an analysis of the Danish web's historical development, b) a description of the needed research infrastructure, including descriptions of the different steps of the analysis, c) an outline of how the project can be implemented in other European countries.*

<http://www.netlab.dk/research/projects/probing-a-nations-web-domain-the-historical-development-of-the-danish-web/>

Quantitative Studies :Big (or fat) data and computationnal tools ...



+ Why do National Web Archives matter for researchers?



- 8 years ago: starting with the Wayback Machine and search by URLs
- An evolving field which moved fast during the research (search by key words or time stamps in the Wayback machine)
- A growing collaboration with French national Web Archives from 2016



BROWSE HISTORY

DONATE

Explore more than 298 billion [web pages](#) saved over time

2017



Explore more than 377 billion [web pages](#) saved over time

DONATE

2019



Tools

- [Wayback Machine Availability API](#)
- [Chrome Extension](#)
- [Firefox Add-on](#)
- [Safari Extension](#)
- [iOS app](#)
- [Android app](#)



Subscription Service

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. [Visit Archive-It to build and browse the collections.](#)

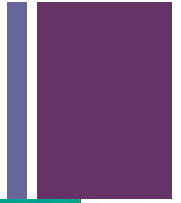


Save Page Now

SAVE PAGE

Capture a web page as it appears now for use as a trusted citation in the future.
[Try New Version! \(beta\)](#)

+ The value of national collections



- More « national » content
- Searchable content at BnF available from 2016

Internet Lab in 2016

Full-text advanced search

{ BnF

(BnF Archives de l'internet Labs) COLLECTIONS ▾ MON COMPTE ▾ [Aide](#) [À propos](#)

[En urgence, collecte sur les attentats parisiens de 2015](#)

Recherche avancée [\(?\)](#)

Mot(s) : Rechercher un ou plusieurs mots ou une "expression exacte"

Proximité : Mot 1 Mot 2 5 Rechercher des mots qui apparaissent à une distance de N

Au moins un de ces deux champs doit être renseigné pour lancer une recherche.

Exclure le(s) mot(s) : Exclure les documents où ce(s) mot(s) apparaissent (facultatif)

Filtres (facultatifs)

Date : Limiter la recherche à une date (aaaa, mm/aaaa ou jj/mm/aaaa)

Ou période : de à Limiter la recherche à une période (aaaa, mm/aaaa ou jj/mm/aaaa)

Hôte, nom de domaine, extension : Limiter la recherche à un hôte (gallica.bnf.fr), un nom de domaine (bnf.fr) ou une extension (fr)

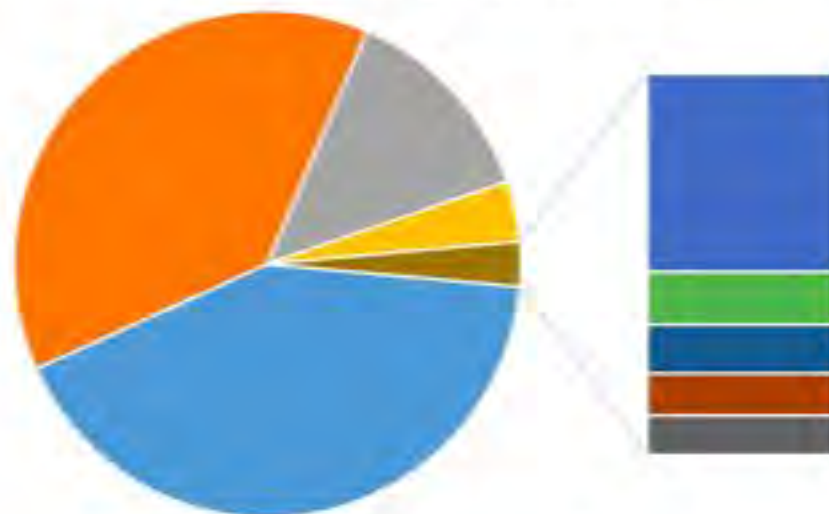
Format : Préciser le format de fichier des documents (html, text, pdf, image, audio, video, word, excel, powerpoint, other)

- Mediations and transparency



Introducing the web-archives: Luxembourg

URLs per language (top 9)



August 2016 crawl

70 million total

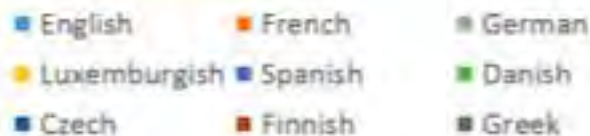
25 million English

..

2 million Luxembourgish

..

9 million others



DH Benelux, Utrecht, 3-7 July 2017

Presentation "Research opportunities for the archived web in the Benelux"

at DH Benelux by Sally Chambers, Peter Mechant, Kees Teszelszky, Yves Maurer

+ The value of national collections

→ Understanding the shaping of Web Archives

→ Dedicated tools

→ Statistics and metadata

→ Sustainability (permalinks, preservation of some corpora, etc.) /

→ In return: common projects

allowing to disseminate the collections' value, meetings to better understand researchers' needs and uses (trading zone)...

→ Next steps: Transnational studies / Engagement in developing literacy (MAHEC Winter School with BnL and Publication Offices of the EU) / Cross-media approaches / Digital Public History ...

